Allowing for probability zero in credal networks under epistemic irrelevance

Jasper De Bock
Ghent University, Belgium
jasper.debock@ugent.be

Gert de Cooman
Ghent University, Belgium
gert.decooman@ugent.be

Abstract

We generalise Cozman's concept of a credal network under epistemic irrelevance [2, Section 8.3] to the case where lower (and upper) probabilities are allowed to be zero. Our main definition is expressed in terms of coherent lower previsions and imposes epistemic irrelevance by means of strong coherence rather than element-wise Bayes's rule. We also present a number of alternative representations for the resulting joint model, both in terms of lower previsions and credal sets, a notable example being an intuitive characterisation of the joint credal set by means of linear constraints. We end by applying our method to a simple case: the independent natural extension for two binary variables. This allows us to, for the first time, find analytical expressions for the extreme points of this special type of independent product.

Keywords. Credal networks, epistemic irrelevance, lower previsions, credal sets, coherence, irrelevant natural extension, independent natural extension.

1 Introduction

Standard Bayesian networks can be generalised to allow for imprecise probability assessments in a multitude of ways; see Ref. [3, Section 3] for an overview. One way to do so is by means of a credal network under epistemic irrelevance. It differs from standard Bayesian networks in two ways: beliefs are modelled by means of closed convex sets of probability measures (so-called *credal sets*) rather than single probability measures, and the non-parent non-descendants of a variable are *epistemically irrelevant* to that variable given its parents, rather than independent of it.

Credal networks under epistemic irrelevance were introduced by Cozman in Ref. [2, Section 8.3]. In order to impose the assessment of epistemic irrelevance, he assumed that all conditioning events have strictly positive lower probability. Under this assumption, a credal set can be conditioned by applying Bayes's rule to each of its probability measures. However, we feel this assumption to be rather restrictive since an event with zero lower probability may have strictly positive upper probability. Therefore, in the present paper, we get rid of this positivity assumption. We do so by using coherent lower previsions as an alternative, equivalent representation for credal sets and using the concept of (strong) coherence to impose epistemic irrelevance assessments, even when the conditioning events have lower or upper probability zero. See Ref. [8] for an earlier successful application of this method to the special case of credal trees.

The graphical structure of a credal network is a directed acyclic graph, of which we recall some basic definitions in Section 2. Section 3 goes on to introduce some basic terminology regarding the variables in the network and we explain in Section 4 how to model a subject's beliefs regarding the values of these variables by means of coherent lower previsions. Section 5 introduces the notion of a credal network under epistemic irrelevance. We first recall how it is defined under the positivity assumption, then provide a definition that does not need that assumption, and prove a number of useful properties and alternative characterisations. We explain how to describe the joint model by means of a set of linear constraints in Section 6, and reformulate this approach in Section 7 for the special case of the so-called independent natural extension. Finally, in Section 8, we apply our method to the independent natural extension of two binary variables and use it to, for the first time, obtain analytical expressions for the extreme points of this extension.

2 Directed acyclic graphs

A directed acyclic graph (DAG) is a graphical model that is well known for its use in Bayesian networks. It consists of a finite set of nodes (vertices), which are joined together into a network by a set of directed edges, each edge connecting one node with another. Since this directed graph is assumed to be acyclic, it is not possible to follow a sequence of directed edges from node to node and end up back at the same node you started out from.

We denote the set of nodes associated with a given DAG by G. For two nodes s and t in G, if there is a directed edge from s to t, we denote this as $s \to t$ and say that s is a *parent* of t and t is a *child* of s. A single node can have multiple parents and multiple children. For any node s, its set of parents is denoted by $P(s)$ and its set of children by $C(s)$. If a node s has no parents, $P(s) = \emptyset$, and we call s a *root node*. If $C(s) = \emptyset$, then we call s a *leaf*, or *terminal node*.

Two nodes s and t, are said to have a *directed path* between them if one can start from s, follow the edges of the DAG taking their direction into account, and end up in t. In other words: one can find a sequence of nodes $s = s_1, \ldots, s_n = t$, $n \geq 1$, in G such that it holds for all $i \in \{1, \ldots, n-1\}$ that $s_i \to s_{i+1}$. In that case we also say that s *precedes* t and write $s \sqsubseteq t$. If $s \sqsubseteq t$ and $s \neq t$, we say that s strictly precedes t and write $s \sqsubset t$. For any node s, we denote its set of *descendants* by $D(s) := \{t \in G : s \sqsubset t\}$, its set of *ascendants* by $A(s) := \{t \in G : t \sqsubset s\}$ and its set of *non-parent non-descendants* by $N(s) := G \setminus (P(s) \cup \{s\} \cup D(s))$.

3 Variables and gambles on them

With each node s in G, we associate a variable X_s taking values in some non-empty finite set $\mathscr{X}_s$. Generic elements of this set are denoted by x_s or z_s. A real-valued function on $\mathscr{X}_s$ is called a gamble and we use $\mathscr{G}(\mathscr{X}_s)$ to denote the set of all of them. Generic gambles are denoted by f, g or γ. As a special kind of gambles we consider *indicators* $\mathbb{I}_A$ of events $A \subseteq \mathscr{X}_s$. $\mathbb{I}_A$ is equal to 1 if the event A occurs (the variable X_s assumes a value in A) and zero otherwise.

We extend this notation to more complicated situations as follows. For any subset S of G, we denote by X_S the tuple of variables (with one component X_s for each $s \in S$) that takes values in the Cartesian product $\mathscr{X}_S := \times_{s \in S} \mathscr{X}_s$. We assume logical independence, meaning that X_S may assume *all* values in $\mathscr{X}_S$. Generic elements of the finite set $\mathscr{X}_S$ are denoted by x_S or z_S. Also, if we mention a tuple x_S, then for any $s \in S$, the corresponding element in the tuple will be denoted by x_s. The set $\mathscr{G}(\mathscr{X}_S)$ contains all gambles on $\mathscr{X}_S$ and $\mathbb{I}_A$ is again used to denote the indicator of an event $A \subseteq \mathscr{X}_S$.

We will frequently use the simplifying device of identifying a gamble f_S on $\mathscr{X}_S$ with its *cylindrical extension* to $\mathscr{X}_U$, where $S \subseteq U \subseteq G$. This is the gamble f_U on $\mathscr{X}_U$ defined by $f_U(x_U) := f_S(x_S)$ for all $x_U \in \mathscr{X}_U$. To give an example, this device allows us to identify the gambles $\mathbb{I}_{\{x_S\}}$ on $\mathscr{X}_S$ and $\mathbb{I}_{\{x_S\} \times \mathscr{X}_{U \setminus S}}$ on $\mathscr{X}_U$, and therefore also the events $\{x_S\}$ and $\{x_S\} \times \mathscr{X}_{U \setminus S}$.

When $S = \emptyset$, we let $\mathscr{X}_\emptyset := \{x_\emptyset\}$ be a singleton. The corresponding variable $X_\emptyset$ can only take this single value $x_\emptyset$, so there is no uncertainty about it. $\mathscr{G}(\mathscr{X}_\emptyset)$ can then be identified with the set $\mathbb{R}$ of real numbers.

4 Modelling beliefs about the network

For two disjoint subsets O and I of G and any $x_I \in \mathscr{X}_I$ we consider two equivalent methods of modelling a subject's beliefs about the value that X_O will assume in $\mathscr{X}_O$, given the observation that $X_I = x_I$.

The first approach is to use a *credal set* $K(X_O|x_I)$, defined as a closed and convex subset of the so-called $\mathscr{X}_O$-simplex $\Sigma_{\mathscr{X}_O}$, which is the set containing all probability mass functions on $\mathscr{X}_O$. A generic element of $K(X_O|x_I)$ is denoted by $p(X_O|x_I)$. It is a probability mass function on $\mathscr{X}_O$ conditional on the observation that $X_I = x_I$

The second approach is to use a *coherent lower prevision* $\underline{P}_O(\cdot|x_I)$, defined as a real-valued functional on $\mathscr{G}(\mathscr{X}_O)$ that satisfies the following three conditions: for all $f, g \in \mathscr{G}(\mathscr{X}_O)$ and all real $\lambda \geq 0$

C1. $\underline{P}_O(f|x_I) \geq \min f$,

C2. $\underline{P}_O(\lambda f|x_I) = \lambda \underline{P}_O(f|x_I)$,

C3. $\underline{P}_O(f+g|x_I) \geq \underline{P}_O(f|x_I) + \underline{P}_O(g|x_I)$.

The conjugate of $\underline{P}_O(\cdot|x_I)$ is called a coherent upper prevision. It is denoted by $\overline{P}_O(\cdot|x_I)$ and defined for all $f \in \mathscr{G}(\mathscr{X}_O)$ by $\overline{P}_O(f|x_I) := -\underline{P}_O(-f|x_I)$. We will focus on coherent lower previsions, but it is useful to keep in mind that all our results can be reformulated in terms of coherent upper previsions by applying this conjugacy property.

Both approaches are equivalent because there is a one-to-one correspondence between them [12, Section 3.3.3]. If we denote by $P_O(\cdot|x_I)$ the expectation operator on $\mathscr{G}(\mathscr{X}_O)$ that corresponds to a probability mass function $p(X_O|x_I)$, then a credal set $K(X_O|x_I)$ defines a unique coherent lower prevision $\underline{P}_O(\cdot|x_I)$ in the following way. For all $f \in \mathscr{G}(\mathscr{X}_O)$:

$$\underline{P}_O(f|x_I) := \min\{P_O(f|x_I) : p(X_O|x_I) \in K(X_O|x_I)\}.$$

Its conjugate coherent upper prevision $\overline{P}_O(\cdot|x_I)$ is given for all $f \in \mathscr{G}(\mathscr{X}_O)$ by

$$\overline{P}_O(f|x_I) := \max\{P_O(f|x_I) : p(X_O|x_I) \in K(X_O|x_I)\}.$$

Conversely, the unique credal set $K(X_O|x_I)$ that corresponds to a coherent lower prevision $\underline{P}_O(\cdot|x_I)$ is given by

$$K(X_O|x_I) := \{p(X_O|x_I) \in \Sigma_{\mathscr{X}_O} : (\forall f \in \mathscr{G}(\mathscr{X}_O)) P_O(f|x_I) \geq \underline{P}_O(f|x_I)\}. \quad (1)$$

If $I = \emptyset$, then $X_I = X_\emptyset$ assumes its only possible value $x_\emptyset$ with certainty, so conditioning on $X_\emptyset = x_\emptyset$ amounts to not conditioning at all. We reflect this in our notation by using $K(X_O)$ and $\underline{P}_O$ as alternative notations for $K(X_O|x_\emptyset)$ and $\underline{P}_O(\cdot|x_\emptyset)$ respectively. A notable example is $I = \emptyset$ and $O = G$, for which we obtain a credal set $K(X_G)$ and coherent lower prevision $\underline{P}_G$ that can be used to model a subject's beliefs about the value that the joint variable X_G will assume in $\mathscr{X}_G$.

When given for all $x_I \in \mathscr{X}_I$, a coherent lower prevision $\underline{P}_O(\cdot|x_I)$ on $\mathscr{G}(\mathscr{X}_O)$, this defines a unique corresponding *coherent conditional lower prevision* $\underline{P}_{O\cup I}(\cdot|X_I)$. It is a special two-place function that is defined, for all $f \in \mathscr{G}(\mathscr{X}_{O\cup I})$ and all $x_I \in \mathscr{X}_I$, by $\underline{P}_{O\cup I}(f|x_I) := \underline{P}_O(f(\cdot,x_I)|x_I)$.

5 Irrelevant natural extension

We will now show how to construct a joint model for the variables in the network in the form of a credal set $K(X_G)$, or equivalently, a coherent lower prevision $\underline{P}_G$.

5.1 Local uncertainty models

We start by adding *local uncertainty models* to each of the nodes $s \in G$. These local models are assumed to be given beforehand and will be used as basic building blocks to construct the joint model.

If s is not a root node of the network, i.e. has a non-empty set of parents $P(s)$, then we have a conditional local model for every instantiation of its parents: for each $x_{P(s)} \in \mathscr{X}_{P(s)}$, we have a credal set $K(X_s|x_{P(s)})$ and a corresponding coherent lower prevision $\underline{P}_s(\cdot|x_{P(s)})$. They represent our subject's beliefs about the variable X_s conditional on the information that its parent variables $X_{P(s)}$ assume the value $x_{P(s)}$.

If s is a root node, i.e. has no parents, then our subject's local beliefs about the variable X_s are represented by an unconditional local model. We are given a credal set $K(X_s)$ and a corresponding coherent lower prevision $\underline{P}_s$. As explained in Section 4, we can also use the common generic notations $K(X_s|x_{P(s)})$ and $\underline{P}_s(\cdot|x_{P(s)})$ in this unconditional case, since for a root node s, its set of parents $P(s)$ is empty.

In order to turn these local uncertainty models into a joint model, we introduce the important concept of epistemic irrelevance.

5.2 Epistemic irrelevance

We discuss conditional epistemic irrelevance, as the unconditional version can easily be recovered as a special case.

Consider three disjoint subsets C, I, and O of G. When a subject judges X_I to be *epistemically irrelevant to X_O conditional on X_C*, he assumes that if he knew the value of X_C, then learning in addition which value X_I assumes in $\mathscr{X}_I$ would not affect his beliefs about X_O. More formally put, he assumes for all $x_C \in \mathscr{X}_C$ and $x_I \in \mathscr{X}_I$ that

$$K(X_O|x_{C\cup I}) = K(X_O|x_C) \text{ and } \underline{P}_O(\cdot|x_{C\cup I}) = \underline{P}_O(\cdot|x_C).$$

It should be clear that it suffices for the unconditional case, in the discussion above, to let $C = \emptyset$. This makes sure the variable X_C has only one possible value, so conditioning on that variable amounts to not conditioning at all.

Using this concept of epistemic irrelevance, we can provide the graphical structure of the network with an interpretation.

5.3 Interpretation of the graphical model

In Bayesian networks, the graphical structure is taken to represent the following assessments: for any node s, the associated variable is independent of its non-parent non-descendant variables, given its parent variables.

When generalising this interpretation to imprecise graphical networks, the classical notion of independence gets replaced by a more general, imprecise-probabilistic notion of independence. In this paper, we choose to use epistemic irrelevance. We provide the graphical structure of the network with the following interpretation: for any node s and all subsets I of its non-parent non-descendants $N(s)$, the variable X_I is judged to be epistemically irrelevant to X_s conditional on $X_{P(s)}$.

More formally put, we assume for all $s \in G$, $I \subseteq N(s)$ and $x_{P(s)\cup I} \in \mathscr{X}_{P(s)\cup I}$ that

$$K(X_s|x_{P(s)\cup I}) := K(X_s|x_{P(s)}) \text{ and } \underline{P}_s(\cdot|x_{P(s)\cup I}) := \underline{P}_s(\cdot|x_{P(s)}).$$

5.4 Non-zero lower probabilities

Together with the local uncertainty models, the irrelevance assessments that are encoded in the network provide us with a number of belief models about the variables in the network: for all $s \in G$, $I \subseteq N(s)$ and $x_{P(s)\cup I} \in \mathscr{X}_{P(s)\cup I}$, we are given a credal set $K(X_s|x_{P(s)\cup I})$, or equivalently, a coherent lower prevision $\underline{P}_s(\cdot|x_{P(s)\cup I})$. In order to arrive at a joint model, we need to provide a method of translating these belief models into constraints on the joint.

An approach that is often used when dealing with assessments of epistemic irrelevance [6, 2], is to assume that all lower probabilities are strictly positive, or equivalently, that for every probability mass function $p(X_G)$ in the joint credal set $K(X_G)$, all events have strictly positive probability. For all $s \in G$, $I \subseteq N(s)$ and $x_{P(s)\cup I} \in \mathscr{X}_{P(s)\cup I}$, this assumption allows us to apply Bayes's rule to every $p(X_G)$ in $K(X_G)$, resulting in a set of conditional probability mass functions $p(X_s|x_{P(s)\cup I})$. This procedure is called applying *elementwise Bayes's rule*. One can now impose that, for all $s \in G$, $I \subseteq N(s)$ and $x_{P(s)\cup I} \in \mathscr{X}_{P(s)\cup I}$, the set of conditional probability mass functions that is obtained in this way must be equal to the given model $K(X_s|x_{P(s)\cup I})$. Any joint credal set $K(X_G)$ that satisfies these constraints is called an *irrelevant product* of the local models.

One particular credal set that was proven to be an irrelevant product in Ref. [2]—under the positivity assumption mentioned above—is the so-called *strong extension* of the network. Its credal set $K^{\text{str}}(X_G)$ is the convex hull of the set $\mathscr{P}$, which contains all joint probability mass functions $p(X_G)$ that, for all $x_G \in \mathscr{X}_G$, satisfy

$$p(x_G) = \prod_{s \in G} p(x_s | x_{P(s)}),$$

where each $p(X_s | x_{P(s)})$ is selected from the local credal set $K(X_s | x_{P(s)})$. The corresponding coherent lower prevision $\underline{P}_G^{\mathrm{str}}$ is given for all $f \in \mathscr{G}(\mathscr{X}_G)$ by

$$\underline{P}_G^{\mathrm{str}}(f) = \min\{P_G(f) : p(X_G) \in \mathscr{P}\}.$$

The strong extension is not the only irrelevant product of the local models. Although it has the advantage of having an intuitive similarity to standard Bayesian networks, it is somewhat arbitrary in that it satisfies more constraints than those needed to be called an irrelevant product. We prefer to use a least committal strategy: to only satisfy those constraints that are imposed by the network, and no others. The resulting model is the largest of all credal sets that are an irrelevant product. We call it the *irrelevant natural extension* of the network an denote it by $K^{\mathrm{irr}}(X_G)$.

This irrelevant natural extension was introduced by Cozman in Ref. [2], but only under the assumption that all lower probabilities are strictly positive. We feel this assumption to be rather restrictive since an event with zero lower probability may occur with a strictly positive upper probability. The first contribution of this paper will therefore be to extend Cozman's definition of the irrelevant natural extension such that it allows for lower (and upper) probabilities to be zero.

5.5 Getting rid of the positivity assumption

If the conditioning event has lower probability zero, the credal set $K(X_s | x_{P(s) \cup I})$ can no longer be uniquely related to the joint model $K(X_G)$ through element-wise Bayes's rule. Therefore, we have to impose our assessments of epistemic irrelevance in some other way. Here, we choose to do so by means of strong coherence, defining the irrelevant natural extension in terms conditional lower previsions, rather than their corresponding credal sets.

As mentioned in the beginning of Section 5.4, the irrelevance assessments, together with the local uncertainty models, provide us with a number of coherent lower previsions: for all $s \in G$, $I \subseteq N(s)$ and $x_{P(s) \cup I} \in \mathscr{X}_{P(s) \cup I}$ we are given a coherent lower prevision $\underline{P}_s(\cdot | x_{P(s) \cup I}) := \underline{P}_s(\cdot | x_{P(s)})$ on $\mathscr{G}(\mathscr{X}_s)$. As was explained in Section 4, this provides us with a number of coherent *conditional* lower previsions: for all $s \in G$ and $I \subseteq N(s)$, we have a coherent conditional lower prevision $\underline{P}_{\{s\} \cup P(s) \cup I}(\cdot | X_{P(s) \cup I})$, defined for all $f \in \mathscr{G}(\mathscr{X}_{\{s\} \cup P(s) \cup I})$ and $x_{P(s) \cup I} \in \mathscr{X}_{P(s) \cup I}$ by

$$\underline{P}_{\{s\} \cup P(s) \cup I}(f | x_{P(s) \cup I}) := \underline{P}_s(f(\cdot, x_{P(s) \cup I}) | x_{P(s)}).$$

We will denote the set consisting of all these conditional lower previsions as $\mathscr{I}(\underline{P}_{\{s\} \cup P(s)}(\cdot | X_{P(s)}), s \in G)$.

In order to turn these coherent conditional lower previsions into constraints on a joint model, given in the form of a coherent lower prevision $\underline{P}_G$ on $\mathscr{G}(\mathscr{X}_G)$, we use the concept of *(strong) coherence* [12, Section 7.1.4]: we require $\underline{P}_G$ to be strongly coherent with the family $\mathscr{I}(\underline{P}_{\{s\} \cup P(s)}(\cdot | X_{P(s)}), s \in G)$ of coherent conditional lower previsions. Any $\underline{P}_G$ that satisfies this property, is called an *irrelevant product*. The least committal—pointwise smallest— irrelevant product is called the *irrelevant natural extension* of the network and will be denoted by $\underline{P}_G^{\mathrm{irr}}$.

As strong coherence is a rather involved requirement, we will not get into the details of what it means. For our present purposes, it suffices to think of it as a generalisation of the element-wise Bayes's rule approach that was explained in Section 5.4. For the interested reader: Ref. [12, Section 7.1.4] provides a general definition and a behavioural interpretation in terms of supremum buying prices, turning strong coherence into a rationality requirement.

We would like to stress that strong coherence is a consistency criterion, rather than a conditioning rule.[1] In fact, it is compatible with a number of fundamentally different conditioning rules, all of which reduce to element-wise Bayes's rule if the conditioning event has positive lower probability. Also, strong coherence regards conditional models as fundamental, rather than deriving them from unconditional ones. In that respect, it shares fundamental ideas with the well-known concept of full conditional measures. See Ref. [1] for a similar, coherence-based approach to stochastic independence, which has been applied to credal networks in Ref. [11].

When it comes to strong coherence, the so-called Reduction Theorem [12, Theorem 7.1.5] is a very useful result; see also Ref. [9, Theorem 2]. It implies that the unconditional coherent lower prevision $\underline{P}_G$ is strongly coherent with the family $\mathscr{I}(\underline{P}_{\{s\} \cup P(s)}(\cdot | X_{P(s)}), s \in G)$ of conditional ones—is an irrelevant product—, if and only if (i) the family $\mathscr{I}(\underline{P}_{\{s\} \cup P(s)}(\cdot | X_{P(s)}), s \in G)$ is strongly coherent on its own and (ii) $\underline{P}_G$ is weakly coherent [12, Section 7.1.4] with $\mathscr{I}(\underline{P}_{\{s\} \cup P(s)}(\cdot | X_{P(s)}), s \in G)$.

Using an approach that uses so-called sets of desirable gambles rather than coherent lower previsions, it is relatively easy to show that requirement (i) is always satisfied [5, Proposition 16].

Proposition 1. *Consider arbitrary coherent lower previsions $\underline{P}_s(\cdot | x_{P(s)})$ on $\mathscr{G}(\mathscr{X}_s)$, $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$. Then the family $\mathscr{I}(\underline{P}_{\{s\} \cup P(s)}(\cdot | X_{P(s)}), s \in G)$ is strongly coherent.*

It follows that $\underline{P}_G$ is an irrelevant product if and only if it

[1] Refs. [7, Definition 12] and [4, Section 3.2.4] provide definitions for epistemic irrelevance that are based on a conditioning rule that is similar to Walley's notion of regular extension [12, Appendix J]. These definitions are applicable in the presence of zero lower probabilities as well. It is not clear to us whether they can be used to construct a joint model from conditional ones, as is done in the current paper.

is weakly coherent with $\mathscr{I}(\underline{P}_{\{s\}\cup P(s)}(\cdot|X_{P(s)}), s \in G)$. In its original form [12, Section 7.1.4], weak coherence is still rather involved, but due to Ref. [9, Theorem 1], it can be reformulated in a very elegant manner that leads directly to the following characterisation of an irrelevant product.

Corollary 2. *A coherent lower prevision $\underline{P}_G$ on $\mathscr{G}(\mathscr{X}_G)$ is strongly coherent with $\mathscr{I}(\underline{P}_{\{s\}\cup P(s)}(\cdot|X_{P(s)}), s \in G)$—is an irrelevant product—if and only if for all $s \in G$, $I \subseteq N(s)$, $x_{P(s)\cup I} \in \mathscr{X}_{P(s)\cup I}$ and $g \in \mathscr{G}(\mathscr{X}_s)$:*

$$\underline{P}_G(\mathbb{I}_{x_{P(s)\cup I}}[g - \underline{P}_s(g|x_{P(s)})]) = 0.$$

The condition imposed in this result is called the *Generalised Bayes's Rule* (GBR), and reduces to element-wise Bayes's rule when all conditioning events have strictly positive lower probabilities [12, Theorem 6.4.2]. It should therefore be clear that the definition of an irrelevant product, as it was given in Section 5.4 under the assumption of strictly positive lower probabilities, is a special case of the definition given in the current section.

Proposition 3. *The strong extension is an irrelevant product: the coherent lower prevision $\underline{P}_G^{\text{str}}$ is strongly coherent with $\mathscr{I}(\underline{P}_{\{s\}\cup P(s)}(\cdot|X_{P(s)}), s \in G)$.*

This result guarantees the existence of at least one irrelevant product, making the irrelevant natural extension well defined: since strong coherence is preserved under taking lower envelopes [12, Section 7.1.6], the irrelevant natural extension is the lower envelope of all irrelevant products, implying that it is indeed pointwise dominated by all other irrelevant products. It should be clear that Corollary 2 provides us with an immediate characterisation for this irrelevant natural extension.

Corollary 4. *The irrelevant natural extension of a network is the pointwise smallest coherent lower prevision $\underline{P}_G^{\text{irr}}$ on $\mathscr{G}(\mathscr{X}_G)$ such that for all $s \in G$, $I \subseteq N(s)$, $x_{P(s)\cup I} \in \mathscr{X}_{P(s)\cup I}$ and $g \in \mathscr{G}(\mathscr{X}_s)$:*

$$\underline{P}_G(\mathbb{I}_{x_{P(s)\cup I}}[g - \underline{P}_s(g|x_{P(s)})]) = 0.$$

Similar to what has been shown in Ref. [2, Lemma 13]—under the positivity assumption—most of the constraints in Corollary 4 turn out to be redundant. We find that we only need to impose those constraints for which $I = N(s)$.

Theorem 5. *The irrelevant natural extension of a network is the pointwise smallest coherent lower prevision $\underline{P}_G^{\text{irr}}$ on $\mathscr{G}(\mathscr{X}_G)$ such that for all $s \in G$, $x_{P(s)\cup N(s)} \in \mathscr{X}_{P(s)\cup N(s)}$ and $g \in \mathscr{G}(\mathscr{X}_s)$:*

$$\underline{P}_G(\mathbb{I}_{x_{P(s)\cup N(s)}}[g - \underline{P}_s(g|x_{P(s)})]) = 0.$$

Although we have defined the irrelevant natural extension in terms of coherent (conditional) lower previsions—since strong coherence is not particularly well-suited for a formulation in terms of credal sets—, it is valid for credal sets as well. Due to the correspondence between credal sets and coherent lower previsions, it suffices to consider the credal set that corresponds to the irrelevant natural extension $\underline{P}_G^{\text{irr}}$. We denote it by $K^{\text{irr}}(X_G)$ and will also refer to it as the irrelevant natural extension of the network. Using Eq. (1), we find that

$$K^{\text{irr}}(X_G) = \{p(X_G) \in \Sigma_{\mathscr{X}_G} : \\ (\forall f \in \mathscr{G}(\mathscr{X}_G)) P_G(f) \geq \underline{P}_G^{\text{irr}}(f)\}.$$

The following result provides an intuitive characterisation.

Theorem 6. *A probability mass function $p(X_G) \in \Sigma_{\mathscr{X}_G}$ belongs to $K^{\text{irr}}(X_G)$ if and only if for all $s \in G$ and $x_{P(s)\cup N(s)} \in \mathscr{X}_{P(s)\cup N(s)}$ there are a real number $\lambda \geq 0$ and a probability mass function $p(X_s|x_{P(s)}) \in K(X_s|x_{P(s)})$ such that*

$$\sum_{z_{D(s)} \in \mathscr{X}_{D(s)}} p(x_{P(s)\cup N(s)}, X_s, z_{D(s)}) = \lambda\, p(X_s|x_{P(s)}).$$

5.6 Marginalisation properties

Given a credal network with nodes G and local models $K(X_s|x_{P(s)})$, $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, a *top sub-network* is a network formed by a subset of nodes $S \subseteq G$ such that for all $s \in S$, its ascendants $A(s)$ also belong to S. The underlying graphical structure consists of those edges in the original network that connect nodes in S and the local models $K(X_s|x_{P(s)})$, $s \in S$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, are taken to be identical to those of the original model. We denote the irrelevant natural extension of such a top sub-network as $K^{\text{irr}}(X_S)$. It turns out to be closely related to the irrelevant natural extension of the original network, a result that was already present in Ref. [2, Theorem 15] under the assumption that all lower probabilities are strictly positive.

Proposition 7. *Consider a credal network with nodes G and a top sub-network with nodes S. Let $K^{\text{irr}}(X_G)$ and $K^{\text{irr}}(X_S)$ be their respective irrelevant natural extensions. Denote by $\text{marg}_S(K^{\text{irr}}(X_G))$ the credal set obtained by element-wise marginalisation to $\mathscr{X}_S$ of the probability mass functions in $K^{\text{irr}}(X_G)$, then*

$$K^{\text{irr}}(X_S) = \text{marg}_S(K^{\text{irr}}(X_G)).$$

We believe that the irrelevant natural extension also satisfies marginalisation properties for sub-networks other than the very specific subclass of top sub-networks, but we defer any formal result to future work. See Ref. [5] to get an idea of what might be possible.

6 A linear programming approach

The goal of the current section is to construct a set of linear constraints that is able to fully characterise the joint credal set $K^{\text{irr}}(X_G)$ of the irrelevant natural extension of a given network.

In order to derive such a representation for the joint model, we start from similar representations for the local models. For all $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, we characterise the local credal set $K(X_s|x_{P(s)})$ as the set of all real-valued functions $p(z_s|x_{P(s)}) \in \mathbb{R}^{\mathscr{X}_s}$ that satisfy the unitary constraint

$$\sum_{z_s \in \mathscr{X}_s} p(z_s|x_{P(s)}) = 1 \tag{2}$$

and a (possibly infinite) set of linear homogeneous inequalities

$$\sum_{z_s \in \mathscr{X}_s} p(z_s|x_{P(s)})\gamma(z_s) \geq 0, \tag{3}$$

where γ takes values in a (possibly infinite) set $\Gamma(s, x_{P(s)})$ of gambles on $\mathscr{X}_s$.

Such a description for $K(X_s|x_{P(s)})$ always exists, as it can be derived from the corresponding coherent lower prevision $\underline{P}_s(\cdot|x_{P(s)})$ by letting

$$\Gamma(s, x_{P(s)}) = \{f - \underline{P}_s(f|x_{P(s)}) : f \in \mathscr{G}(\mathscr{X}_s)\}. \tag{4}$$

Indeed, for this particular choice of $\Gamma(s, x_{P(s)})$, the combination of Eqs. (2) and (3) will always be equivalent with the constraints imposed by Eq. (1), thereby fully characterising $K(X_s|x_{P(s)})$. To understand why this equivalence holds, start by noticing that if $\gamma = f - \underline{P}_s(f|x_{P(s)})$, with $f \in \mathscr{G}(\mathscr{X}_s)$, then due to Eq. (2), Eq. (3) becomes equivalent to

$$\sum_{z_s \in \mathscr{X}_s} p(z_s|x_{P(s)})f(z_s) \geq \underline{P}_s(f|x_{P(s)}). \tag{5}$$

Coherence of $\underline{P}_s(\cdot|x_{P(s)})$ now implies, for all $z_s \in \mathscr{X}_s$, that $\underline{P}_s(\mathbb{I}_{\{z_s\}}|x_{P(s)}) \geq 0$ and therefore, due to Eq. (5), that $p(z_s|x_{P(s)}) \geq 0$. By combining this with Eq. (2), we find that $p(X_s|x_{P(s)}) \in \Sigma_{\mathscr{X}_s}$. This allows us to rewrite the left-hand side of Eq. (5) as $P_s(f|x_{P(s)})$, thereby establishing the equivalence with the constraints imposed by Eq. (1).

Eq. (4) produces an infinite set of constraints that is guaranteed to characterise $K(X_s|x_{P(s)})$, but in practice, most of these constraints will often be redundant. This is especially the case for so-called *finitely generated* local models, for which the corresponding coherent lower prevision $\underline{P}_s(\cdot|x_{P(s)})$ is fully determined by its value in only a finite number of gambles. For such local models, one can easily construct a set $\Gamma(s, x_{P(s)})$ that contains only a finite number of constraints and yet fully characterises $K(X_s|x_{P(s)})$. The credal set of such a finitely generated local model will always be the convex hull of a finite number of probability mass functions. The reason for this equivalence being that a compact convex set can be specified as the intersection of a finite number of closed half spaces if and only if it is the convex hull of a finite number of vertices [10, Theorem 3.1.3].

The importance of these local representations in terms of linear constraints—regardless of whether $\Gamma(s, x_{P(s)})$ is finite or not—is that we can use the local constraints to derive global ones, thereby obtaining the following representation for the irrelevant natural extension of a network.

Proposition 8. *Consider a credal network for which each of the local credal sets $K(X_s|x_{P(s)})$, $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, is fully characterised by means of Eqs.* (2) *and* (3)*. Then $K^{\text{irr}}(X_G)$ consists of those $p(X_G) \in \Sigma_{\mathscr{X}_G}$ for which for all $s \in G$, $x_{P(s)\cup N(s)} \in \mathscr{X}_{P(s)\cup N(s)}$ and $\gamma \in \Gamma(s, x_{P(s)})$:*

$$\sum_{z_s \in \mathscr{X}_s} \sum_{z_{D(s)} \in \mathscr{X}_{D(s)}} p(x_{P(s)\cup N(s)}, z_s, z_{D(s)})\gamma(z_s) \geq 0.$$

When all lower probabilities are strictly positive, this result is fairly straightforward. The global inequalities can then be obtained by imposing all irrelevancies through element-wise Bayes's rule and clearing the denominators, as is done in Ref. [2, Section 8.3]. The importance of our result is that it shows that these inequalities remain valid if lower (and upper) probabilities are allowed to be zero.

Ref. [2] does not explicitly impose $p(X_G) \in \Sigma_{\mathscr{X}_G}$ as a constraint. It seems to assume that it suffices to impose only the unitary constraint $\sum_{z_G \in \mathscr{X}_G} p(z_G) = 1$, making the requirement that $p(z_G) \geq 0$, $z_G \in \mathscr{X}_G$, redundant. Although we agree with this statement, we do not believe it to be trivial and therefore choose to provide it with a proof.

Theorem 9. *Consider a credal network for which each of the local credal sets $K(X_s|x_{P(s)})$, $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, is fully characterised by means of Eqs.* (2) *and* (3)*. Then $K^{\text{irr}}(X_G)$ consists of those real-valued functions $p(X_G) \in \mathbb{R}^{\mathscr{X}_G}$ for which $\sum_{z_G \in \mathscr{X}_G} p(z_G) = 1$ and for all $s \in G$, $x_{P(s)\cup N(s)} \in \mathscr{X}_{P(s)\cup N(s)}$ and $\gamma \in \Gamma(s, x_{P(s)})$:*

$$\sum_{z_s \in \mathscr{X}_s} \sum_{z_{D(s)} \in \mathscr{X}_{D(s)}} p(x_{P(s)\cup N(s)}, z_s, z_{D(s)})\gamma(z_s) \geq 0.$$

Proposition 8 and Theorem 9 are valid for both finite and infinite sets $\Gamma(s, x_{P(s)})$, but in the infinite case, their value is mainly of a theoretical nature. They can only be used in practice—at least in an exact way—if $L(s, x_{P(s)})$ is finite for all $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, or equivalently, if all local credal sets are finitely generated.[2] Indeed, in that case, Proposition 8 and Theorem 9 will provide linear programs with a finite number of constraints. Although the size of these programs is still exponential in the number of variables that define the network, it allows for inference problems in small networks to be solved in an exact manner. Initial ideas on how to reduce this exponential complexity are provided in our conclusions.

7 Independent natural extension

An important special case is obtained when all nodes in the network are unconnected. Every node $s \in G$ is then both

[2] If we allow for non-linear constraints, then local credal sets that are not finitely generated could be practical as well, as they can often be described by means of a finite set of non-linear constraints. We believe that Proposition 8 and Theorem 9 could be adapted easily to allow for such non-linear (homogeneous) constraints, thereby expanding their practical use when combined with non-linear solvers.

a root and a leaf of the network—meaning that $P(s)$ and $C(s)$ are empty—, its non-parent non-descendants are given by $N(s) = G \setminus \{s\}$ and the local model is an unconditional credal set $K(X_s)$, or equivalently, a coherent lower prevision $\underline{P}_s$ on $\mathscr{G}(\mathscr{X}_s)$.

For such a network, the irrelevancies that are encoded by the network are the following. For every $s \in G$ and all $I \subseteq G \setminus \{s\}$, the variable X_I is epistemically irrelevant to X_s, implying that for any two nodes $s,t \in G$, X_s and X_t are mutually epistemically irrelevant and therefore by definition *epistemically independent*. The resulting irrelevant natural extension is called the *many-to-one independent natural extension* and has been treated in full detail in Ref. [9]. That same reference also introduces the so-called *many-to-many independent natural extension*, which requires that for all disjoint subsets O and I of G, X_I is epistemically irrelevant to X_O. The many-to-one and many-to-many independent natural extensions are shown to be equivalent [9, Theorem 23] and we can therefore simply call it the *independent natural extension*. Its coherent lower prevision is denoted by $\otimes_{s \in G} \underline{P}_s$ and its credal set by $\otimes_{s \in G} K(X_s)$. For this special case, Theorem 9 can be reformulated in the following way.

Corollary 10. *Consider a finite number of local credal sets $K(X_s)$, $s \in G$, each of which is fully characterised means of Eqs.* (2) *and* (3)*. Then $\otimes_{s \in G} K(X_s)$ consists of those real-valued functions $p(X_G) \in \mathbb{R}^{\mathscr{X}_G}$ for which $\sum_{z_G \in \mathscr{X}_G} p(z_G) = 1$ and for all $s \in G$, $x_{G \setminus \{s\}} \in \mathscr{X}_{G \setminus \{s\}}$ and $\gamma \in \Gamma(s)$:*

$$\sum_{z_s \in \mathscr{X}_s} p(x_{G \setminus \{s\}}, z_s) \gamma(z_s) \geq 0.$$

We leave it to the reader to reformulate some of the other results that were obtained in the two previous sections, taking the simplifications that correspond to the special case of the independent natural extension into account. In fact, Ref. [9, Proposition 14, Corollary 16 and Theorem 20] already provides results that could be regarded as special cases of Proposition 3, Corollary 2 and Proposition 7.

8 Case study of two binary variables

As an example, we apply our results to the very simple case of two unconnected binary variables X_1 and X_2. For all $i \in \{1,2\}$, the variable X_i assumes values in its binary state space $\mathscr{X}_i = \{h_i, t_i\}$ and has a given local uncertainty model in the form of a credal set $K(X_i)$. We set out to construct the independent natural extension $K(X_1) \otimes K(X_2)$ of these two local models. In order to do so, we will describe it by means of linear constraints and then use this characterisation to find analytical expressions for the so-called *extreme points* of $K(X_1) \otimes K(X_2)$, which are those elements of $K(X_1) \otimes K(X_2)$ that cannot be written as a convex combination of the other elements. $K(X_1) \otimes K(X_2)$ is then equal to the convex hull of these extreme points.

For a binary variable X_i, $i \in \{1,2\}$, the credal set $K(X_i)$ is uniquely characterised by the lower and upper probability of h_i, respectively denoted as $\underline{p}(h_i)$ and $\overline{p}(h_i)$. Each of these two probabilities defines a mass function on $\mathscr{X}_i$ and

$$K(X_i) = \left\{ p \in \Sigma_{\mathscr{X}_i} : p(h_i) \in [\underline{p}(h_i), \overline{p}(h_i)] \right\}$$

is obtained by taking their convex hull. The corresponding lower and upper probability of t_i is given by $\underline{p}(t_i) := 1 - \overline{p}(h_i)$ and $\overline{p}(t_i) := 1 - \underline{p}(h_i)$.

In order to apply the method described in Section 6, we first need to characterise $K(X_i)$ by means of the unitary constraint and a finite number of linear homogeneous inequalities. In this particular binary case, the following two inequalities suffice:

$$\begin{aligned} \overline{p}(t_i)p(h_i) - \underline{p}(h_i)p(t_i) &\geq 0 \\ -\underline{p}(t_i)p(h_i) + \overline{p}(h_i)p(t_i) &\geq 0. \end{aligned}$$

By applying Corollary 10, these local inequalities can be used to obtain eight global inequalities.

$$\overline{p}(t_1)p(h_1,h_2) - \underline{p}(h_1)p(t_1,h_2) \geq 0 \tag{I1}$$
$$-\underline{p}(t_1)p(h_1,h_2) + \overline{p}(h_1)p(t_1,h_2) \geq 0 \tag{I2}$$
$$\overline{p}(t_1)p(h_1,t_2) - \underline{p}(h_1)p(t_1,t_2) \geq 0 \tag{I3}$$
$$-\underline{p}(t_1)p(h_1,t_2) + \overline{p}(h_1)p(t_1,t_2) \geq 0 \tag{I4}$$
$$\overline{p}(t_2)p(h_1,h_2) - \underline{p}(h_2)p(h_1,t_2) \geq 0 \tag{I5}$$
$$-\underline{p}(t_2)p(h_1,h_2) + \overline{p}(h_2)p(h_1,t_2) \geq 0 \tag{I6}$$
$$\overline{p}(t_2)p(t_1,h_2) - \underline{p}(h_2)p(t_1,t_2) \geq 0 \tag{I7}$$
$$-\underline{p}(t_2)p(t_1,h_2) + \overline{p}(h_2)p(t_1,t_2) \geq 0 \tag{I8}$$

Together with the global unitary constraint

$$p(h_1,h_2) + p(h_1,t_2) + p(t_1,h_2) + p(t_1,t_2) = 1,$$

they fully characterise the credal set $K(X_1) \otimes K(X_2)$. If the inequalities in equations (I1)–(I8) are replaced by equalities, we refer to them as (E1)–(E8).

Lemma 11. *Every extreme point of $K(X_1) \otimes K(X_2)$ is the unique solution to the unitary constraint and three of the equations* (E1)–(E8).

The extreme points of the independent natural extension $K(X_1) \otimes K(X_2)$ can therefore be found in the following way. We need to consider every possible subset of three equalities out of (E1)–(E8). For every such combination of three equalities, we need to combine them with the unitary constraint and check whether this results in a unique solution, and if so, whether this unique solution satisfies the inequalities in (I1)–(I8). If so, that unique solution is an extreme point of $K(X_1) \otimes K(X_2)$.

As there are 56 possible ways of choosing three equalities out of eight, one might suspect that this problem cannot be

	$p(h_1,h_2)\Sigma$	$p(h_1,t_2)\Sigma$	$p(t_1,h_2)\Sigma$	$p(t_1,t_2)\Sigma$	Σ
p_{S1}	$\underline{p}(h_1)\underline{p}(h_2)$	$\underline{p}(h_1)\overline{p}(t_2)$	$\overline{p}(t_1)\underline{p}(h_2)$	$\overline{p}(t_1)\overline{p}(t_2)$	1
p_{S2}	$\underline{p}(h_1)\overline{p}(h_2)$	$\underline{p}(h_1)\underline{p}(t_2)$	$\overline{p}(t_1)\overline{p}(h_2)$	$\overline{p}(t_1)\underline{p}(t_2)$	1
p_{S3}	$\overline{p}(h_1)\underline{p}(h_2)$	$\overline{p}(h_1)\overline{p}(t_2)$	$\underline{p}(t_1)\underline{p}(h_2)$	$\underline{p}(t_1)\overline{p}(t_2)$	1
p_{S4}	$\overline{p}(h_1)\overline{p}(h_2)$	$\overline{p}(h_1)\underline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(h_2)$	$\underline{p}(t_1)\underline{p}(t_2)$	1
p_{A1}	$\underline{p}(h_1)\overline{p}(h_1)\underline{p}(h_2)$	$\underline{p}(h_1)\overline{p}(h_1)\overline{p}(t_2)$	$\overline{p}(t_1)\overline{p}(h_1)\underline{p}(h_2)$	$\underline{p}(h_1)\underline{p}(t_1)\overline{p}(t_2)$	$\underline{p}(h_1)\overline{p}(t_2)+\overline{p}(h_1)\underline{p}(h_2)$
p_{A2}	$\underline{p}(h_1)\overline{p}(h_1)\overline{p}(h_2)$	$\underline{p}(h_1)\overline{p}(h_1)\underline{p}(t_2)$	$\underline{p}(h_1)\underline{p}(t_1)\overline{p}(h_2)$	$\overline{p}(t_1)\overline{p}(h_1)\underline{p}(t_2)$	$\underline{p}(h_1)\overline{p}(h_2)+\overline{p}(h_1)\underline{p}(t_2)$
p_{A3}	$\overline{p}(h_1)\overline{p}(t_1)\underline{p}(h_2)$	$\underline{p}(t_1)\underline{p}(h_1)\overline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(t_1)\underline{p}(h_2)$	$\underline{p}(t_1)\overline{p}(t_1)\overline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(t_2)+\overline{p}(t_1)\underline{p}(h_2)$
p_{A4}	$\underline{p}(t_1)\underline{p}(h_1)\overline{p}(h_2)$	$\overline{p}(h_1)\overline{p}(t_1)\underline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(t_1)\overline{p}(h_2)$	$\underline{p}(t_1)\overline{p}(t_1)\underline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(h_2)+\overline{p}(t_1)\underline{p}(t_2)$
p_{B1}	$\underline{p}(h_2)\overline{p}(h_2)\underline{p}(h_1)$	$\overline{p}(t_2)\overline{p}(h_2)\underline{p}(h_1)$	$\underline{p}(h_2)\overline{p}(h_2)\overline{p}(t_1)$	$\underline{p}(h_2)\underline{p}(t_2)\overline{p}(t_1)$	$\underline{p}(h_2)\overline{p}(t_1)+\overline{p}(h_2)\underline{p}(h_1)$
p_{B2}	$\overline{p}(h_2)\overline{p}(t_2)\underline{p}(h_1)$	$\underline{p}(t_2)\overline{p}(t_2)\underline{p}(h_1)$	$\underline{p}(t_2)\underline{p}(h_2)\overline{p}(t_1)$	$\underline{p}(t_2)\overline{p}(t_2)\overline{p}(t_1)$	$\underline{p}(t_2)\overline{p}(t_1)+\overline{p}(t_2)\underline{p}(h_1)$
p_{B3}	$\underline{p}(h_2)\overline{p}(h_2)\overline{p}(h_1)$	$\underline{p}(h_2)\underline{p}(t_2)\overline{p}(h_1)$	$\underline{p}(h_2)\overline{p}(h_2)\underline{p}(t_1)$	$\overline{p}(t_2)\overline{p}(h_2)\underline{p}(t_1)$	$\underline{p}(h_2)\overline{p}(h_1)+\overline{p}(h_2)\underline{p}(t_1)$
p_{B4}	$\underline{p}(t_2)\underline{p}(h_2)\overline{p}(h_1)$	$\underline{p}(t_2)\overline{p}(t_2)\overline{p}(h_1)$	$\overline{p}(h_2)\overline{p}(t_2)\underline{p}(t_1)$	$\underline{p}(t_2)\overline{p}(t_2)\underline{p}(t_1)$	$\underline{p}(t_2)\overline{p}(h_1)+\overline{p}(t_2)\underline{p}(t_1)$

Table 1: Candidates for the extreme points of the independent natural extension of two binary variables

solved manually. However, due to the extreme symmetry—switching X_1 and X_2, h_1 and t_1 or h_2 and t_2 yields an equivalent set of inequalities—, only 7 of those 56 cases need to be considered, as the others can be related to these 7 by an argument of symmetry. In this way, we managed to obtain analytical expressions for the extreme points of $K(X_1) \otimes K(X_2)$.

Theorem 12. *Analytical expressions for the extreme points of $K(X_1) \otimes K(X_2)$ can be found by means of Table 1 and Figure 1. Table 1 contains expressions for 12 probability mass functions, which can be obtained by dividing the numbers in columns 2–5 by the denominator in column 6. The diagram in Figure 1 shows, depending on the particular values of $\underline{p}(h_1)$, $\overline{p}(h_1)$, $\underline{p}(t_1)$, $\overline{p}(t_1)$, $\underline{p}(h_2)$, $\overline{p}(h_2)$, $\underline{p}(t_2)$ and $\overline{p}(t_2)$, which of these 12 probability mass functions are extreme points of $K(X_1) \otimes K(X_2)$. In this diagram, we use the shorthand notation $p_{\mathrm{S1=S2}}$ to denote that p_{S1} and p_{S2} are two coinciding extreme points.*

Although the diagram in Figure 1 considers quite a number of special or degenerate cases, the main result can be summarised quite easily. If one of the local models is precise or vacuous, then the independent natural extension has the same extreme points as—and therefore coincides with—the strong extension. In all other cases, the independent natural extension has up to four additional extreme points.

9 Summary and Conclusions

In this paper, we have developed a definition for credal networks under epistemic irrelevance that allows for zero lower (and upper) probabilities, generalising Cozman's definition [2, Section 8.3], which requires the lower probabilities of conditioning events to be strictly positive. For the resulting joint model, we have derived a number of properties and alternative characterisations. Some of these results were already mentioned by Cozman, but are now proved to remain valid when his positivity requirement is dropped. One particular result is that the joint credal set that corresponds to a credal network under epistemic irrelevance can be described by means of linear constraints. As a first toy example, we have used this approach to obtain analytical expressions for the extreme points of the independent natural extension of two binary variables.

The main future goal that we intend to pursue is to develop algorithms for credal networks under epistemic irrelevance that are able to perform inference in an efficient manner. This problem has been tackled before by Cozman [2, Section 8.4], but we suspect that a more efficient solution can be obtained. The idea would be to derive counterparts to the marginalisation and graphoid properties that are proven in Ref. [5] and combine these with a linear programming approach that builds upon Theorem 9.

Acknowledgements

Jasper De Bock is a Ph.D. Fellow of the Fund for Scientific Research – Flanders (FWO) and wishes to acknowledge the financial support of the FWO. The authors also wish to thank three anonymous referees for their helpful comments.

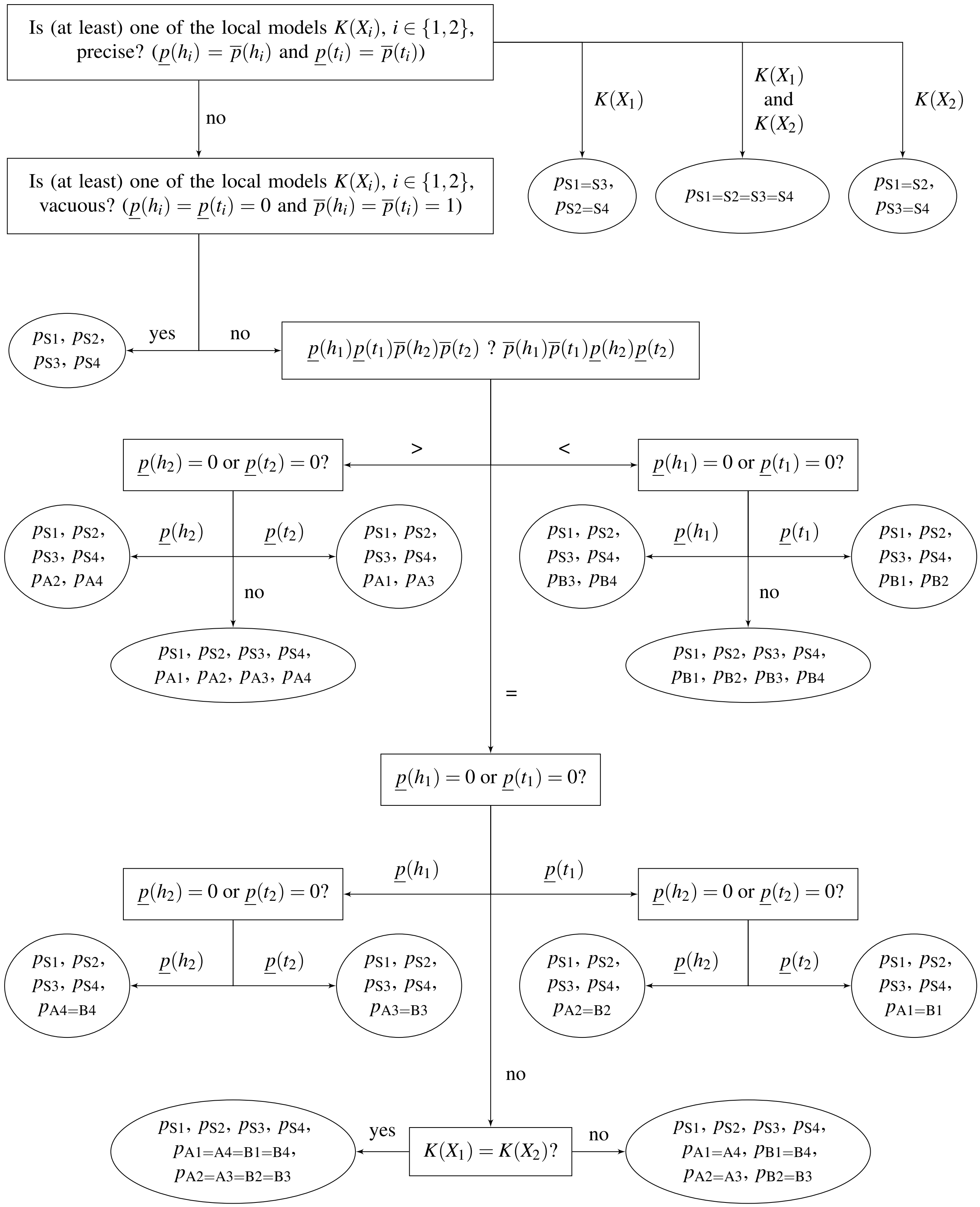

Figure 1: Diagram to obtain the extreme points of the independent natural extension of two binary variables

References

[1] G. Coletti and R. Scozzafava. Stochastic independence in a coherent setting. *Annals of Mathematics and Artificial Intelligence*, 35(1-4):151–176, 2002.

[2] F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[3] F. G. Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39(2-3):167–184, 2005.

[4] F. G. Cozman. Sets of probability distributions, independence, and convexity. *Synthese*, 186(2):577–600, 2012.

[5] J. De Bock and G. de Cooman. Credal networks under epistemic irrelevance using sets of desirable gambles. Accepted for publication in *Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications*, 2013.

[6] C. P. de Campos and F. G. Cozman. Computing lower and upper expectations under epistemic independence. *International Journal of Approximate Reasoning*, 44(3):244–260, 2007.

[7] L. M. De Campos and S. Moral. Independence concepts for convex sets of probabilities. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, UAI'95, pages 108–115, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.

[8] G. de Cooman, F. Hermans, A. Antonucci, and M. Zaffalon. Epistemic irrelevance in credal nets: the case of imprecise Markov trees. *International Journal of Approximate Reasoning*, 51:1029–1052, 2010.

[9] G. de Cooman, E. Miranda, and M. Zaffalon. Independent natural extension. *Artificial Intelligence*, 175:1911–1950, 2011.

[10] B. Grünbaum. *Convex polytopes*. Springer, 2nd edition, prepared by V. Kaibel, V. Klee, and G. M. Ziegler, 2003.

[11] B. Vantaggi. Conditional independence structures and graphical models. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(05):545–571, 2003.

[12] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.